DEPARTMENT OF THE ARMY
U.S. Army Corps of Engineers
CECW-EG   Washington, DC 20314-1000   ETL 1110-2-547

Technical Letter
No. 1110-2-547   30 September 1997

Engineering and Design
INTRODUCTION TO PROBABILITY AND RELIABILITY METHODS
FOR USE IN GEOTECHNICAL ENGINEERING

Distribution Restriction Statement

Approved for public release; distribution is unlimited.

Report Documentation Page

Report Date 30 Sep 1997	Report Type N/A	Dates Covered (from... to) -

Title and Subtitle		Contract Number
Engineering and Design: Introduction to Probability and Reliability Methods for Use in Geotechnical Engineering		Grant Number
		Program Element Number

Author(s)		Project Number
		Task Number
		Work Unit Number

Performing Organization Name(s) and Address(es)	Performing Organization Report Number
Department of the Army U.S. Army Corps of Engineers Washington, DC 20314-1000	

Sponsoring/Monitoring Agency Name(s) and Address(es)	Sponsor/Monitor's Acronym(s)
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes

Abstract

Subject Terms

Report Classification unclassified	Classification of this page unclassified
Classification of Abstract unclassified	Limitation of Abstract UU

Number of Pages
14

Engineering and Design
INTRODUCTION TO PROBABILITY AND RELIABILITY METHODS
FOR USE IN GEOTECHNICAL ENGINEERING

1. Purpose

This engineer technical letter (ETL) provides an introduction to the use of probabilistic methods in geotechnical engineering.

2. Applicability

This ETL applies to HQUSACE elements and USACE commands having responsibility for the design of civil works projects.

3. References

See Appendix A.

4. Discussion

This is the first in a series of ETL's that will provide guidance on the use and application of probability and reliability methods of analyses for use in the assessment of existing levees for benefit determination and the geotechnical portion of major rehabilitation reports.

FOR THE DIRECTOR OF CIVIL WORKS:

APP A - References
APP B - Introduction to Probability and
        Reliability in Geotechnical
        Engineering

DOUGLAS J. KAMIEN
Acting Chief, Engineering Division
Directorate of Civil Works

APPENDIX A: REFERENCES

a.   The following references are U.S. Army Corps of Engineers project-related references.

(1) ETL 1110-2-321, Reliability Assessment of Navigation Structures: Stability of Existing Gravity Structures.

(2) ETL 1110-2-532, Reliability Assessment of Navigation Structures.

(3) Shannon and Wilson, Inc., and Wolff, T. F. 1994. "Probability Models for Geotechnical Aspects of Navigation Structures," report to the St. Louis District, U.S. Army Corps of Engineers.

(4) Wolff, Thomas F. 1994. "Evaluating the Reliability of Existing Levees," report to U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

(5) Wolff, T. F., and Wang, W. 1992. "Engineering Reliability of Navigation Structures," research report, Michigan State University, for U.S. Army Corps of Engineers.

(6) Wolff, T. F., and Wang, W. 1992. "Engineering Reliability of Navigation Structures—Supplement No. 1," research report, Michigan State University, for U.S. Army Corps of Engineers.

b.   The following are those other than U.S. Army Corps of Engineers project-related references.

(1) Ang, A. H.-S., and Tang, W. H. 1975. "Volume I: Basic Principles," *Probability Concepts in Engineering Planning and Design*, John Wiley and Sons, New York.

(2) Ang, A. H.-S., and Tang, W. H. 1984. "Volume II: Decision, Risk, and Reliability," *Probability Concepts in Engineering Planning and Design*, John Wiley and Sons, New York.

(3) Benjamin, J., and Cornell, C. A. 1970. *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York.

(4) Harr, M. E. 1987. *Reliability Based Design in Civil Engineering*, McGraw-Hill, New York.

(5) Hasofer, A. A., and Lind, A. M. 1974. "An Exact and Invariant Second-Moment Code Format," *Journal of the Engineering Mechanics Division*, *ASCE*, 100, 111-121.

(6) Peter, P. 1982. *Canal and River Levees*, Elsevier Scientific Publishing Company, Amsterdam.

(7) Rosenblueth, E. 1975. "Point Estimates for Probability Moments," *Proceedings of the National Academy of Science,* USA, 72(10).

(8) Rosenblueth, E. 1981. "Two-Point Estimates in Probabilities," *Applied Mathematical Modeling*, 5.

(9) Vrouwenvelder, A. C. W. M. 1987. "Probabilistic Design of Flood Defenses," Report No. B-87-404, IBBC-TNO (Institute for Building Materials and Structures of The Netherlands Organization for Applied Scientific Research), The Netherlands.

APPENDIX B: INTRODUCTION TO PROBABILITY AND RELIABILITY IN GEOTECHNICAL ENGINEERING

B-1. Introduction

a. The objective of this ETL is to introduce some basic elements of engineering reliability analysis applicable to geotechnical structures for various modes of performance. These reliability measures are intended to be sufficiently consistent and suitable for application to economic analysis of geotechnical structures of water resource projects. References are provided which should be consulted for detailed discussion of the principles of reliability analyses.

b. Traditionally, evaluations of geotechnical adequacy are expressed by safety factors. A safety factor can be expressed as the ratio of capacity to demand. The safety concept, however, has short-comings as a measure of the relative reliability of geotechnical structures for different performance modes. A primary deficiency is that parameters (material properties, strengths, loads, etc.) must be assigned single, precise values when the appropriate values may in fact be uncertain. The use of precisely defined single values in an analysis is known as the *deterministic* approach. The safety factor using this approach reflects the condition of the feature, the engineer's judgment, and the degree of conservatism incorporated into the parameter values.

c. Another approach, the *probabilistic* approach, extends the safety factor concept to explicitly incor-porate uncertainty in the parameters. This uncertainty can be quantified through statistical analysis of existing data or judgmentally assigned. Even if judgmentally assigned, the probabilistic results will be more meaningful than a deterministic analysis because the engineer provides a measure of the uncertainty of his or her judgment in each parameter.

B-2. Reliability Analysis Principles

a. The probability of failure.

(1) Engineering reliability analysis is concerned with finding the *reliability R* or the *probability of failure Pr(f)* of a feature, structure, or system. As a system is considered reliable unless it fails, the reliability and probability of failure sum to unity:

$$R + Pr(f) = 1$$

$$R = 1 - Pr(f)$$

$$Pr(f) = 1 - R$$

(2) In the engineering reliability literature, the term *failure* is used to refer to any occurrence of an adverse event under consideration, including simple events such as maintenance items. To distinguish adverse but noncatastrophic events (which may require repairs and associated expenditures) from events of catastrophic failure (as used in the dam safety context), the term *probability of unsatisfactory performance Pr(U)* is sometimes used. An example would be slope stability where the safety factor is below the required minimum safety factor but above 1.0. Thus, for this case, reliability is defined as:

$$R = 1 - Pr(U)$$

b. Contexts of reliability analysis.

(1) Engineering reliability analysis can be used in several general contexts:

- The estimation of the reliability of a new structure or system upon its construction and first loading.

- The estimation of the reliability of an existing structure or system upon a new loading.

- The estimation of the probability of a part or system surviving for a given lifetime.

Note that the third context has an associated time interval, where as the first two involve measures of the overall adequacy of the system in response to a load event.

(2) Reliability for the first two contexts can be calculated using the *capacity-demand model* and quantified by the *reliability index* β.   In the capacity-demand model, uncertainty in the performance of the structure or system is taken to be a function of the

uncertainty in the values of various parameters used in calculating some measure of performance, such as the factor of safety.

(3) In the third context, reliability over a future time interval is calculated using parameters developed from actual data on the lifetimes or frequencies of failure of similar parts or systems. These are usually taken to follow the exponential or Weibull probability distributions. This methodology is well established in electrical, mechanical, and aerospace engineering where parts and components routinely require periodic replacement. This approach produces a *hazard function* which defines the probability of failure in any time period. These functions are used in economic analysis of proposed geotechnical improvements. The development of hazard functions is not part of this ETL.

(4) For reliability evaluation of most geotechnical structures, in particular existing levees, the capacity-demand model will be utilized, as the question of interest is the probability of failure related to a load event rather than the probability of failure within a time interval.

c. Reliability index. The reliability index β is a measure of the reliability of an engineering system that reflects both the mechanics of the problem and the uncertainty in the input variables. This index was developed by the structural engineering profession to provide a measure of comparative reliability without having to assume or determine the shape of the probability distribution necessary to calculate an exact value of the probability of failure. The reliability index is defined in terms of the expected value and standard deviation of the performance function, and permits comparison of reliability among different structures or modes of performance without having to calculate absolute probability values. Calculating the reliability index requires:

- A deterministic model (e.g., a slope stability analysis procedure).

- A performance function (e.g., the factor of safety from UTEXAS2).

- The expected values and standard deviations of the parameters taken as random variables (e.g., $E[\phi]$ and σ_ϕ).

- A definition of the limit state (e.g., ln *(FS)* = 0).

- A method to estimate the expected value and standard deviation of the limit state given the expected values and standard deviations of the parameters (e.g., the Taylor's series or point estimate methods).

d. Accuracy of reliability index.

(1) For rehabilitation studies of geotechnical structures, the reliability index is used as a "relative measure of reliability or confidence in the ability of a structure to perform its function in a satisfactory manner."

(2) The analysis methods used to calculate the reliability index should be sufficiently accurate to rank the relative reliability of various structures and components. However, reliability index values are not absolute measures of probability. Structures, components, and performance modes with higher indices are considered more reliable than those with lower indices. Experience analyzing geotechnical structures will refine these techniques.

B-3. The Capacity-Demand Model

a. In the capacity-demand model, the probability of failure or unsatisfactory performance is defined as the probability that the demand on a system or component exceeds the capacity of the system or component. The capacity and demand can be combined into a single function (*the performance function)*, and the event that the capacity equals the demand taken as the *limit state.* The *reliability R* is the probability that the limit state will not be achieved or crossed.

b. The concept of the capacity-demand model is illustrated for slope stability analysis in Figure B-1. Using the expected value and standard deviation of the random variables *c* and φ in conjunction with the Taylor's series method or the point estimate method, the expected value and standard deviation of the factor of safety can be calculated. If it is assumed that the factor of safety is lognormally distributed, then the natural log of the factor of safety is normally distributed. The performance function is taken as the log of
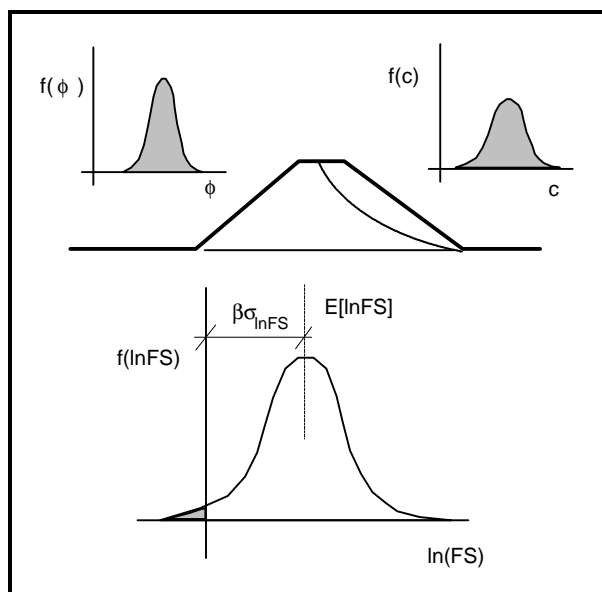
Figure B-1. The capacity-demand model

variable in an analysis can be modeled as a random variable as most properties and parameters have some inherent variability and uncertainty. However, a few specific random variables will usually dominate the analysis. Including additional random variables may unnecessarily increase computational effort without significantly improving results. When in doubt, a few analyses with and without certain random variables will quickly illustrate which are significant, as will the examination of variance terms in a Taylor's series analysis. For levee analysis, significant random variables typically include material strengths, soil permeability or permeability ratio, and thickness of top stratum. Material properties such as soil density may be significant, but where strength and density both appear in an analysis, strength may dominate. An example of a variable that can be represented deterministically (nonrandom) is the density of water.

the factor of safety, and the limit state is taken as the condition ln $(FS) = 0$. The probability of failure is then the shaded area corresponding to the condition ln $(FS) < 0$. If it is assumed that the distribution on ln (FS) is normal, then the probability of failure can be obtained using standard statistical tables.

c. Equivalent performance functions and limit states can be defined using other measures, such as the exit gradient for seepage.

d. The probability of failure associated with the reliability index is a *probability per structure*; it has no time-frequency basis. Once a structure is constructed or loaded as modeled, it either performs satisfactorily or not. Nevertheless, the β value calculated for an existing structure provides a rational comparative measure.

B-4. Steps in a Reliability Analysis Using the Capacity-Demand Model

As suggested by Figure B-1 for slope stability, a reliability analysis includes the following steps:

- Important variables considered to have sufficient inherent uncertainty are taken as random variables and characterized by their expected values, standard deviations, and correlation coefficients. In concept, every

- A performance function and limit state are identified.

- The expected value and standard deviation of the performance function are calculated. In concept, this involves integrating the performance function over the probability density functions of the random variables. In practice, approximate values are obtained using the expected value, standard deviation, and correlation coefficients of the random variables in the Taylor's series method or the point estimate method.

- The reliability index β is calculated from the expected and standard deviation of the performance function. The reliability index is a measure of the distance between the expected value of ln (C/D) or ln (FS) and the limit state.

- If a probability of failure value is desired, a distribution is assumed and *Pr(f)* calculated.

B-5. Random Variables

a. Description. Parameters having significance in the analysis and some significant uncertainty are taken as *random variables.* Instead of having precise single values, random variables assume a range of values in accordance with a *probability density*

B-3

function or *probability distribution.* The probability distribution quantifies the likelihood that its value lies in any given interval. Two commonly used distributions, the normal and the lognormal, are described later in this appendix.

   b. Moments of random variables. To model random variables in the Taylor's series or point estimate methods, one must provide values of their expected values and standard deviations, which are two of several probabilistic *moments* of a random variable. These can be calculated from data or estimated from experience. For random variables which are not independent of each other, but tend to vary together, correlation coefficients must also be assigned.

   (1) Mean value. The *mean* value μ_X of a set of N measured values for the random variable X is obtained by summing the values and dividing by N:

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$$

   (2) Expected value. The *expected value E[X]* of a random variable is the mean value one would obtain if all possible values of the random variable were multiplied by their likelihood of occurrence and summed. Where a mean value can be calculated from representative data, it provides an unbiased estimate of the expected value of a parameter; hence, the mean and expected value are numerically the same. The expected value is defined as:

$$E[X] = \mu_X = \int X f(X)\, dx \approx \sum X p(X_i)$$

where

   $f(X) =$ *probability density function* of X (for continuous random variables)
   $p(X_i) =$ probability of the value X_i (for discrete random variables)

   (3) Variance. The *variance Var [X]* of a random variable X is the expected value of the squared difference between the random variable and its mean value. Where actual data are available, the variance of the data can be calculated by subtracting each value from the mean, squaring the result, and determining the average of these values:

$$Var[X] = E\left[(X - \mu_X)^2\right] = \int (X - \mu_X)^2 f(X)\, dX$$

$$= \frac{\sum \left[(X_i - \mu_X)^2\right]}{N}$$

The summation form above involving the X_i term provides the variance of a population containing exactly N elements. Usually, a *sample* of size N is used to obtain an *estimate of the variance* of the associated random variable which represents an *entire population* of items or continuum of material. To obtain an unbiased estimate of the population working from a finite sample, the N is replaced by $N - 1$:

$$Var[X] = \frac{\sum \left[(X_i - \mu_X)^2\right]}{N - 1}$$

   (4) Standard deviation. To express the scatter or dispersion of a random variable about its expected value in the same units as the random variable itself, the *standard deviation* σ_X is taken as the square root of the variance; thus:

$$\sigma_x = \sqrt{Var[X]}$$

   (5) Coefficient of variation. To provide a convenient dimensionless expression of the uncertainty inherent in a random variable, the standard deviation is divided by the expected value to obtain the *coefficient of variation* V_X which is usually expressed as a percent:

$$V_X = \frac{\sigma_X}{E[X]} \times 100 \ \%$$

The expected value, standard deviation, and coefficient of variation are interdependent: knowing any two, the third is known. In practice, a convenient way to estimate moments for parameters where little data are available is to assume that the coefficient of variation is similar to previously measured values from other data sets for the same parameter.

   c. Correlation. Pairs of random variables may be correlated or independent; if correlated, the likelihood of a certain value of the random variable Y depends on the value of the random variable X. For example, the strength of sand may be correlated with density or the top blanket permeability may be correlated with grain size of the sand. The *covariance Cov [X, Y]* is analogous to the variance but measures

the combined effect of how two variables vary together. The definition of the covariance is:

$$Cov[X,Y] = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

which is equivalent to:

$$Cov[X,Y] = \int \int (X - \mu_X)(Y - \mu_Y) f(X,Y) \; dY \; dX$$

In the above equation, *f(X,Y)* is the joint probability density function of the random variables *X* and *Y*. To calculate the covariance from data, the following equation can be used:

$$Cov[X,Y] = \frac{1}{N} \sum (X_i - \mu_X)(Y_i - \mu_Y)$$

To provide a nondimensional measure of the degree of correlation between *X* and *Y*, the *correlation coefficient* $\rho_{X,Y}$ is obtained by dividing the covariance by the product of the standard deviations:

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

The correlation coefficient may assume values from -1.0 to +1.0. A value of 1.0 or -1.0 indicates there is perfect linear correlation; given a value of *X*, the value of *Y* is known and hence is not random. A value of zero indicates no linear correlation between variables. A positive value indicates the variables increase and decrease together; a negative value indicates that one variable decreases as the other increases. Pairs of *independent* random variables have zero correlation coefficients.

B-6. Probability Distributions

   a.   Definition.

   (1) The terms *probability distribution* and *probability density function pdf* or the notation $f_X(X)$ refer to a function that defines a continuous random variable. The Taylor's series and point estimate methods described herein to determine moments of performance functions require only the mean and standard deviation of random variables and their correlation coefficients; knowledge of the form of the probability density function is not necessary. However, in order to ensure that estimates made for these moments are reasonable, it is recommended that the engineer plot the shape of

the normal or lognormal distribution which has the expected value and standard deviation assumed. This can easily be done with spreadsheet software.

   (2) Figure B-1 illustrated probability density functions for the random variables *c* and ϕ. A probability density function has the property that for any *X*, the value of *f(x)* is proportional to the like-lihood of *X*. The area under a probability density function is unity. The probability that the random variable *X* lies between two values X_1 and X_2 is the integral of the probability density function taken between the two values. Hence:

$$Pr(X_1 < X < X_2) = \int\limits_{X_1}^{X_2} f_X(X) \; dx$$

   (3) The *cumulative distribution function CDF* or $F_X(X)$ measures the integral of the probability density function from minus infinity to *X*:

$$F_X(X) = \int\limits_{-\infty}^{X} f_X(X) \; dx$$

Thus for any value *X*, $F_X(X)$ is the probability that the random variable *X* is less than the given *x*.

   b.   Estimating probabilistic distributions. A suggested method to assign or check assumed moments for random variables is to:

- Assume trial values for the expected value and standard deviation and take the random variable to be normal or lognormal.

- Plot the resulting density function and tabulate and plot the resulting cumulative distribution function (spreadsheet software is a convenient way to do this).

- Assess the reasonableness of the shape of the *pdf* and the values of the *CDF*.

- Repeat above steps with successively improved estimates of the expected value and standard deviation until an appropriate *pdf* and *CDF* are obtained.

   c.   Normal distribution. The *normal* or *Gaussian* distribution is the most well-known and widely

assumed probability density function. It is defined in terms of the mean μ_X and standard deviation σ_X as:

$$f_X(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$$

When fitting the normal distribution, the mean of the distribution is taken as the expected value of the random variable. The cumulative distribution function for the normal distribution is not conveniently expressed in closed form but is widely tabulated and can be readily computed by numerical approximation. It is a built-in function in most spreadsheet programs. Although the normal distribution has limits of plus and minus infinity, values more than three or four standard deviations from the mean have very low probability. Hence, one empirical fitting method is to take minimum and maximum reasonable values to be at plus and minus three or so standard deviations. The normal distribution is commonly assumed to characterize many random variables where the coefficient of variation is less than about 30 percent. For levees, these include soil density and drained friction angle. Where the mean and standard deviation are the only information known, it can be shown that the normal distribution is the most unbiased choice.

    d. Lognormal distribution.

    (1) When a random variable X is lognormally distributed, its natural logarithm, ln X, is normally distributed. The lognormal distribution has several properties which often favor its selection to model certain random variables in engineering analysis:

- As X is positive for any value of ln X, lognormally distributed random variables cannot assume values below zero.

- It often provides a reasonable shape in cases where the coefficient of variation is large (> 30 percent) or the random variable may assume values over one or more orders of magnitude.

- The central limit theorem implies that the distribution of products or ratios of random variables approaches the lognormal distribution as the number of random variables increases.

    (2) If the random variable X is lognormally distributed, then the random variable $Y = \ln X$ is normally distributed with parameters $E[Y] = E[\ln X]$

and $\sigma_Y = \sigma_{\ln X}$. To obtain the parameters of the normal random variable Y, first the coefficient of variation of X is calculated:

$$V_X = \frac{\sigma_X}{E[X]}$$

The standard deviation of Y is then calculated as:

$$\sigma_Y = \sigma_{\ln X} = \sqrt{\ln\left(1 + V_X^2\right)}$$

The standard deviation σ_Y is in turn used to calculate the expected value of Y:

$$E[Y] = E[\ln X] = \ln E[X] - \frac{\sigma_Y^2}{2}$$

The density function of the lognormal variate X is:

$$f(X) = \frac{1}{X\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln X - E[Y]^2}{\sigma_Y}\right)\right]$$

The shape of the distribution can be plotted from the above equation. Values on the cumulative distribution function for X can be determined from the cumulative distribution function of Y (E[Y], σ_Y) by substituting the X in the expression $Y = \ln X$.

B-7. Calculation of the Reliability Index

    a. Figure B-2 illustrates that a simple definition of the reliability index is based on the assumption that capacity and demand are normally distributed and the limit state is the event that their difference, the safety margin S, is zero. The random variable S is then also normally distributed and the reliability index is the distance by which $E[S]$ exceeds zero in units of σ_S:

$$\beta = \frac{E[S]}{\sigma_S} = \frac{E[C - D]}{\sqrt{\sigma_C^2 + \sigma_D^2}}$$

An alternative formulation (also shown in Figure B-2) implies that capacity C and demand D are lognormally distributed random variables. In this case ln C and ln D are normally distributed. Defining the factor of safety FS as the ratio C/D, then ln $FS = (\ln C) - (\ln D)$ and ln FS is normally distributed. Defining the reliability index as the distance by which ln FS
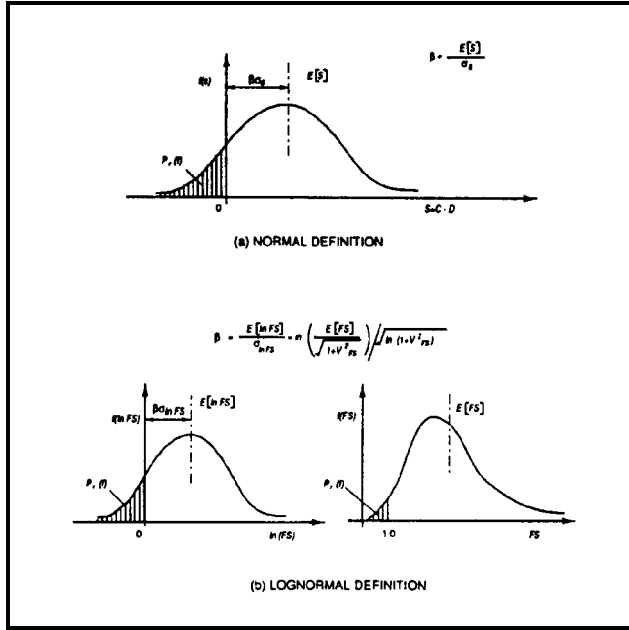
Figure B-2. Alternative definitions of the reliability index

exceeds zero in terms of the standard deviation of ln *FS*, it is:

$$\beta = \frac{E[\ln C - \ln D]}{\sigma_{(\ln C - \ln D)}} = \frac{E[\ln (C/D)]}{\sigma_{\ln(C/D)}}$$

$$= \frac{E[\ln FS]}{\sigma_{\ln FS}}$$

b. From the properties of the lognormal distribution, the expected value of ln *C* is:

$$E[\ln C] = \ln E[C] - \frac{1}{2}\sigma^2_{\ln C}$$

where:

$$\sigma^2_{\ln C} = \ln\left[1 + V^2_C\right]$$

Similar expressions apply to $E[\ln D]$ and $\sigma_{\ln D}$.

c. The expected value of the log of the factor of safety is then:

$$E[\ln FS] = \ln E[C] - \ln E[D] - \frac{1}{2}\ln\left[1 + V^2_C\right]$$

$$+ \frac{1}{2}\ln\left[1 + V^2_D\right]$$

As the second-order terms are small when the coefficients of variation are not exceedingly large (below approximately 30 percent), the equation above is sometimes approximated as:

$$E[\ln FS] \approx \ln E[C] - \ln E[D] = \ln\left(\frac{E[C]}{E[D]}\right)$$

The standard deviation of the log of the factor of safety is obtained as:

$$\sigma_{\ln FS} = \sqrt{\sigma^2_{\ln C} + \sigma^2_{\ln D}}$$

$$\sigma_{\ln FS} = \sqrt{\ln\left[1 + V^2_C\right] + \ln\left[1 + V^2_D\right]}$$

Introducing an approximation,

$$\ln\left[1 + V^2_C\right] = \approx V^2_C$$

the reliability index for lognormally distributed *C, D,* and *FS* and normally distributed ln *C,* ln *D,* and ln *FS* can be expressed approximately as:

$$\beta = \frac{\ln\left(\frac{E[C]}{E[D]}\right)}{\sqrt{V^2_C + V^2_D}}$$

The exact expression is:

$$\beta = \ln\left(\frac{E[C]\sqrt{1 + V^2_D}}{E[D]\sqrt{1 + V^2_C}}\right)$$

For many geotechnical problems and related deterministic computer programs, the output is in the form of the factor of safety, and the capacity and demand are not explicitly separated. The reliability index must be calculated from values of *E[FS]* and σ_{FS} obtained from multiple runs as described in the next section. In this case, the reliability index is obtained using the following steps:

$$V_{FS} = \frac{\sigma_{FS}}{E[FS]}$$

$$\sigma_{\ln FS} = \sqrt{\ln\left(1 + V^2_{FS}\right)}$$

$$E[\ln FS] = \ln E[FS] - \frac{1}{2}\ln\left(1 + V^2_{FS}\right)$$

B-7

$$\beta = \frac{E[\ln FS]}{\sigma_{\ln FS}} = \frac{\ln\left(E[FS] \bigg/ \sqrt{1 + V_{FS}^2}\right)}{\sqrt{\ln\left(1 + V_{FS}^2\right)}}$$

B-8. Integration of the Performance Function

Methods such as direct integration, Taylor's series method, point estimate method, and Monte Carlo simulation are available for calculating the mean and standard deviation of the performance function. For direct integration, the mean value of the function is obtained by integrating over the probability density function of the random variables. A brief description of the other methods follows. The references should be consulted for additional information.

a.   Taylor's series method. Taylor's series method is one of several methods to estimate the moments of a performance function based on moments of the input random variables. It is based on a Taylor's series expansion of the performance function about some point. For the Corps' navigation rehabilitation studies, the expansion is performed about the expected values of the random variables. The Taylor's series method is termed a first-order, second-moment (FOSM) method as only first-order (linear) terms of the series are retained and only the first two moments (mean and the standard deviation) are considered. The method is summarized below.

(1) Independent random variables. Given a function $Y = g(X_1, X_2, \dots X_n)$, where all X_i values are independent, the expected value of the function is obtained by evaluating the function at the expected values of the random variables:

$$E[Y] = g\left(E[X_1], E[X_2], \dots E[X_n]\right)$$

For a function such as the factor of safety, this implies that the expected value of the factor of safety is calculated using the expected values of the random variables:

$$E[FS] = FS\left(E[\phi_1], E[c_1], E[\gamma_1], \dots\right)$$

The variance of the performance function is taken as:

$$Var[Y] = \sum \left[\left(\frac{\partial Y}{\partial X_i}\right)^2 Var\, X_i\right]$$

with the partial derivatives taken at the expansion point (in this case the mean or expected value). Using the factor of safety as an example performance function, the variance is obtained by finding the partial derivative of the factor of safety with respect to each random variable evaluated at the expected value of that variable, squaring it, multiplying it by the variance of that random variable, and summing these terms over all of the random variables:

$$Var[FS] = \sum \left[\left(\frac{\partial FS}{\partial X_i}\right)^2 Var\, X_i\right]$$

The standard deviation of the factor of safety is then simply the square root of the variance.

(a) Having the expected value and variance of the factor of safety, the reliability index can be calculated as described earlier in this appendix. Advantages of the Taylor's series method include the following:

- The relative magnitudes of the terms in the above summation provide an explicit indication of the relative contribution of uncertainty of each variable.

- The method is exact for linear performance functions.

Disadvantages of the Taylor's series method include the following:

- It is necessary to determine the value of derivatives.

- The neglect of higher order terms introduces errors for nonlinear functions.

(b) The required derivatives can be estimated numerically by evaluating the performance function at two points. The function is evaluated at one increment above and below the expected value of the random variable X_i and the difference of the results is divided by the difference between the two values of X_i. Although the derivative at a point is most precisely evaluated using a very small increment, evaluating the derivative over a range of plus and minus one standard deviation may better capture some of the nonlinear behavior of the function over a range of likely values. Thus, the derivative is evaluated using the following approximation:

$$\frac{\partial Y}{\partial X_i} = \frac{g\left(E[X_i] + \sigma_{X_i}\right) - g\left(E[X_i] - \sigma_{X_i}\right)}{2\sigma_{X_i}}$$

When the above expression is squared and multiplied by the variance, the standard deviation term in the denominator cancels the variance, leading to

$$\left(\frac{\partial Y}{\partial X_i}\right)^2 Var\,X = \frac{[g(X_+) - g(X_-)]^2}{2}$$

where X_+ and X_- are values of the random variable at plus and minus one standard deviation from the expected value.

(2) Correlated random variables. Where random variables are correlated, the solution is more complex. The expression for the expected value, retaining second-order terms is:

$$E[Y] = g\left[E[X_1], E[X_2], ...E[X_n]\right.$$

$$\left. + \frac{1}{2}\sum \frac{\partial^2 Y}{\partial X_i \partial X_j}\,Cov(X_i X_j)\right]$$

However, in keeping with the first-order approach, the second-order terms are generally neglected, and the expected value is calculated the same as for independent random variables. The variance, however, is taken as:

$$Var[Y] = \sum \left[\left(\frac{\partial Y}{\partial X_i}\right)^2 Var\,X_i\right]$$

$$+ 2\sum \left[\frac{\partial Y}{\partial X_i}\frac{\partial Y}{\partial X_j}\,Cov(X_i X_j)\right]$$

where the covariance term contains terms for each possible combination of random variables.

b. Point estimate method. An alternative method to estimate moments of a performance function based on moments of the random variables is the *point estimate method.* Point estimate methods are procedures where probability distributions for continuous random variables are modeled by discrete "equivalent" distributions having two or more values. The elements of these discrete distributions (or *point estimates*) have specific values with defined probabilities such that the first few moments of the discrete distribution match

that of the continuous random variable. Having only a few values over which to integrate, the moments of the performance function are easily obtained. A simple and straightforward point estimate method has been proposed by Rosenblueth (1975, 1981) and is summarized by Harr (1987). That method is briefly summarized below.

(1) Independent random variables. As shown in Figure B-3, a continuous random variable *X* is represented by two point estimates, X_+ and X_-, with probability concentrations P_+ and P_-, respectively. As the two point estimates and their probability concentrations form an equivalent probability distribution for the random variable, the two *P* values must sum to unity. The two point estimates and probability concentrations are chosen to match three moments of the random variable. When these conditions are satisfied for symmetrically distributed random variables, the point estimates are taken at the mean plus or minus one standard deviation:

$$X_{i+} = E[X_i] + \sigma_{X_i}$$
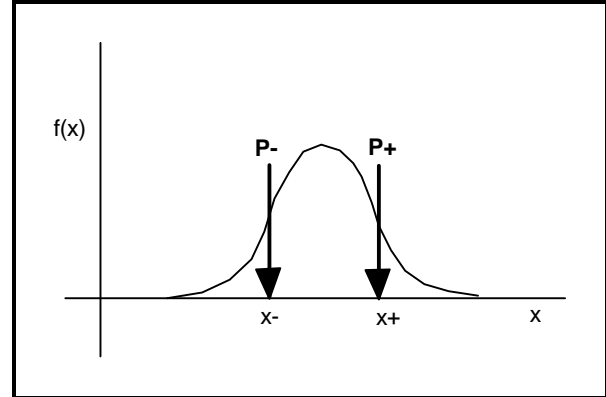
$$X_{i-} = E[X_i] - \sigma_{X_i}$$

Figure B-3. Point estimate method

For independent random variables, the associated probability concentrations are each one-half:

$$P_{i+} = P_{i-} = 0.50$$

Knowing the point estimates and their probability concentrations for each variable, the expected value of a function of the random variables raised to any power *M* can be approximated by evaluating the function for each possible combination of the point estimates (e.g., X_{1+}, X_{2-}, X_{3+}, X_{n-}), multiplying each

result by the product of the associated probability concentrations (e.g., $P_{+--} = P_{1+} P_{2-} P_{3-}$) and summing the terms. For example, two random variables result in four combinations of point estimates and four terms:

$$E[Y^M] = P_{++} g(X_{1+}, X_{2+})^M + P_{+-} g(X_{1+}, X_{2-})^M$$
$$+ P_{-+} g(X_{1-}, X_{2+})^M + P_{--} g(X_{1-}, X_{2-})^M$$

For N random variables, there are 2^N combinations of the point estimates and 2^N terms in the summation. To obtain the expected value of the performance function, the function $g(X_1, X_2)$ is calculated 2^N times using all the combinations and the exponent M is 1. To obtain the standard deviation of the performance function, the exponent M is taken as 2 and the squares of the obtained results are weighted and summed to obtain $E[Y^2]$. The variance can then be obtained from the identity

$$Var[Y] = E[Y^2] - (E[Y])^2$$

and the standard deviation is the square root of the variance.

(2) Correlated random variables. Correlation between symmetrically distributed random variables is treated by adjusting the probability concentrations $(P \pm \pm \ldots \pm)$. A detailed discussion is provided by Rosenblueth (1975) and summarized by Harr (1987). For certain geotechnical analyses involving lateral earth pressure, bearing capacity of shallow foundations, and slope stability, often only two random variables (c and ϕ or $\tan \phi$) need to considered as correlated. For two correlated random variables within a group of two or more, the product of their concentrations is modified by adding a correlation term:

$$P_{i+j-} = P_{i-j+} = (P_{i-})(P_{j+}) - \frac{\rho}{4}$$

$$P_{i+j+} = P_{i-j-} = (P_{i+})(P_{j+}) + \frac{\rho}{4}$$

c. Monte Carlo simulation. The performance function is evaluated for many possible values of the random variables. A plot of the results will produce an approximation of the probability distribution. Once the probability distribution is determined in this manner, the mean and standard deviation of the distribution can be calculated.

B-9. Determining the Probability of Failure

Once the expected value and standard deviation of the performance function have been determined using the Taylor's series or point estimate methods, the reliability index can be calculated as previously described. If the reliability index is assumed to be the number of standard deviations by which the expected value of a normally distributed performance function (e.g., ln (*FS*)) exceeds zero, than the probability of failure can be calculated as:

$$Pr(f) = \psi(-\beta) = \psi(-z)$$

where $\psi(-z)$ is the cumulative distribution function of the standard normal distribution evaluated at *-z*, which is widely tabulated and available as a built-in function on modern microcomputer spreadsheet programs.

B-10. Overall System Reliability

Reliability indices for a number of components or a number of modes of performance may be used to estimate the overall reliability of an embankment. There are two types of systems that bound the possible cases, the series system and the parallel system.

a. Series system. In a series system, the system will perfrom unsatisfactorily if any one component performs unsatisfactorily. If a system has *n* components in series, the probability of unsatisfactory performance of the *ith* component is p_i and its reliability, $R_i = 1 - p_i$, then the reliability of the system, or probability that all components will perform satisfactorily, is the product of the component reliabilities.

$$R = R_1 R_2 R_3 ... R_i ... R_n$$
$$= (1 - p_1)(1 - p_2)(1 - p_3)...(1 - p_i)...(1 - p_n$$

b. Simple parallel system. In a parallel system, the system will only perform unsatisfactorily if all components perform unsatisfactorily. Thus, the reliability is unity minus the probability that all components perform unsatisfactorily, or

$$R = 1 - p_1 p_2 p_3 ... p_i ... p_n$$

c. Parallel series systems.

(1) Solutions are available for systems requiring *r*-out-of-*n* operable components, which may be applicable to problems such as dewatering with multiple pumps, where *r* is defined as the number of reliable units. Subsystems involving independent parallel and series systems can be mathematically combined by standard techniques.

(2) Upper and lower bounds on system reliability can be determined by considering all components to be from subgroups of parallel and series systems, respectively; however, the resulting bounds may be so broad as to be impractical. A number of procedures are found in the references to narrow the bounds.

(3) Engineering systems such as embankments are complex and have many performance modes. Some of these modes may not be independent; for instance, several performance modes may be correlated to the occurrence of a high or low pool level. Rational estimation of the overall reliability of an embankment is a topic that is beyond this ETL.

d. A practical approach.

(1) The reliability of a few subsystems or components may govern the reliability of the entire system. Thus, developing a means to characterize and compare the reliability of these components as a function of time is sufficient to make engineering judgments to prioritize operations and maintenance expenditures.

(2) For initial use in reliability assessment of geotechnical systems, the target reliability values presented below should be used. The objective of the Major Rehabilitation Program would be to keep the reliability index for each significant mode above the target value for the foreseeable future.

B-11. Target Reliability Indices

Reliability indices are a relative measure of the current condition and provide a qualitative estimate of the expected performance. Embankments with relatively high reliability indices will be expected to perform their function well. Embankments with low reliability indices will be expected to perform poorly and present major rehabilitation problems. If the reliability indices are very low, the embankment may be classified as a hazard. The target reliability values shown in Table B-1 should be used in general.

Table B-1. Target Reliability Indices

Expected Performance Level	Beta	Probability of Unsatisfactory Performance
High	5	0.0000003
Good	4	0.00003
Above average	3	0.001
Below average	2.5	0.006
Poor	2.0	0.023
Unsatisfactory	1.5	0.07
Hazardous	1.0	0.16

Note: Probability of unsatisfactory performance is the probability that the value of performance function will approach the limit state, or that an unsatisfactory event will occur. For example, if the performance function is defined in terms of slope instability, and the probability of unsatisfactory performance is 0.023, then 23 of every 1,000 instabilities will result in damage which causes a safety hazard.